A new entropy power inequality for integer-valued random variables

Saeid Haghighatshoar, Emmanuel Abbe, Emre Telatar

Emails: {saeid.haghighatshoar@epfl.ch, eabbe@princeton.edu, emre.telatar@epfl.ch}

Abstract—The entropy power inequality (EPI) provides lower bounds on the differential entropy of the sum of two independent real-valued random variables in terms of the individual entropies. Versions of the EPI for discrete random variables have been obtained for special families of distributions with the differential entropy replaced by the discrete entropy, but no universal inequality is known (beyond trivial ones). More recently, the sumset theory for the entropy function provides a sharp inequality $H(X + X') - H(X) \geq \frac{1}{2} - o(1)$ when X, X' are i.i.d. with high entropy. This paper provides the inequality $H(X + X') - H(X) \geq g(H(X))$, where X, X' are arbitrary i.i.d. integer-valued random variables and where g is a universal strictly positive function on $\mathbb{R}_+$ satisfying $g(0) = 0$. Extensions to non identically distributed random variables and to conditional entropies are also obtained.

Index Terms—Entropy inequalities, Entropy power inequality, Mrs. Gerber's lemma, Doubling constant, Shannon sumset theory.

I. INTRODUCTION

For a continuous random variable[1] $\boldsymbol{X}$ on $\mathbb{R}^n$, let $h(\boldsymbol{X})$ be the differential entropy of $\boldsymbol{X}$ and let $N(\boldsymbol{X}) = 2^{\frac{2}{n}h(\boldsymbol{X})}$ denote the *entropy power* of $\boldsymbol{X}$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two i.i.d. continuous random variables over $\mathbb{R}^n$, the EPI states that

$$N(\boldsymbol{X} + \boldsymbol{Y}) \geq N(\boldsymbol{X}) + N(\boldsymbol{Y}), \tag{1}$$

with equality if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are Gaussian with proportional covariance matrices. A weaker statement of the EPI, yet of key use in applications, is the following inequality stated here for $n = 1$,

$$h(X + X') - h(X) \geq \frac{1}{2}, \tag{2}$$

where X, X' are i.i.d., and where equality holds if and only if X is Gaussian.

The EPI was first proposed by Shannon [1] who used a variational argument to show that Gaussian random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ with proportional covariance matrices and specified differential entropies constitute a stationary point for $h(\boldsymbol{X} + \boldsymbol{Y})$. However, this does not exclude saddle points and local minima. The first rigorous proof of the EPI was given by Stam [2] in 1959, using the De Bruijin's identity which connects the derivative of the entropy with Gaussian perturbation to the Fisher information. This proof was further simplified by Blachman [3]. Another proof was proposed by Lieb [4] based on an extension of Young's inequality.

---

[1] All continuous random variables are assumed to have well-defined differential entropies.

While there is a wide range of inequalities involving union of random variables, the EPI is the only general inequality in information theory estimating the entropy of a sum of independent random variables by means of the individual entropies. It is used as a key ingredient to prove converse results in coding theorems [8]–[12].

There have been numerous extensions and simplifications of the EPI over the reals [6], [7], [13]–[21]. There have also been several attempts to obtain discrete versions of the EPI, using Shannon entropy. Of course, it is not clear what is meant by a discrete version of the EPI, since (1), (2) clearly do no hold verbatim for Shannon entropy.

Several extensions have yet been developed. First, there have been some extensions using finite field additions, for example, the so-called Mrs. Gerber's Lemma (MGL) proved in [23] by Wyner and Ziv for binary alphabets. The MGL was further extended by Witsenhausen [24] to non binary alphabets, who also provided counter-examples for the case of general alphabets. More recently, [28] obtained EPI and MGL results for abelian groups of order 2^n. Second, concerning discrete random variables and addition over the reals, Harremoes and Vignat [25] proved that the discrete EPI holds for binomial random variables with parameter $\frac{1}{2}$, which later on was generalized by Sharma, Das and Muthukrishnan [26]. Yu and Johnson [27] obtained a version of the EPI for discrete random variables using the notion of thinning.

More recently, Tao established in [29] a sumset theory for Shannon entropy, obtaining in particular the sharp inequality

$$H(X + X') - H(X) \geq \frac{1}{2} - o(1),$$

where $o(1)$ vanishes when $H(X)$ tends to infinity. Further results were obtained for the differential entropy in [30].

In this paper, we are interested in integer-valued random variables with arithmetic over the reals. We show that there exists an increasing function $g : \mathbb{R}_+ \to \mathbb{R}_+$, such that $g(x) = 0$ if and only if $x = 0$, and

$$H(X + X') - H(X) \geq g(H(X)),$$

for any i.i.d. integer-valued random variables X, X'. Although we have provided an explicit characterization of g, we found that proving the existence of such a function (even without explicit characterization) is equally challenging. We further generalize the result to non identically distributed random variables and to conditional entropies. We also discuss some open problems in Section IV, in particular, a closure convexity

conjecture which would strengthen the conditional entropy result.

The results obtained in this paper were used in [22] to prove a polarization coding result for discrete random variables using Hadamard matrices over the reals.

Notation: The set of integers and reals will be denoted by $\mathbb{Z}$ and $\mathbb{R}$. Similarly, $\mathbb{Z}_+$ and $\mathbb{R}_+$ will denote the set of positive integers and positive reals. We will use large letters for random variables and small letters for their realizations (the random variable X can have realization x). The natural logarithm and the logarithm in base 2 will be denoted by $\ln$ and $\log_2$ respectively and for $x \in [0,1]$, $h_2(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ will denote the binary entropy function with the convention that $0\log_2(0) = 0$. The entropy of a discrete random variable X in base 2 (bits) will be denoted by $H(X)$. We will interchangeably use $H(p)$ or $H(X)$, where p is the probability distribution of X. The conditional entropy of a random variable X given another random variable Y will be denoted by $H(X|Y)$. For $a, b \in \mathbb{R}$, we will use $a \vee b$ and $a \wedge b$ for the maximum and minimum of a and b. Also $a^+ = a \vee 0$ will denote the positive part of a.

II. RESULTS

In this section, we will give an overview of the results proved in the paper. The first theorem gives a lower bound on the entropy gap of sum of two i.i.d. random variables as a function of their entropies.

Theorem 1. *There is a function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any two i.i.d. $\mathbb{Z}$-valued random variables X, X' with probability distribution p,*

$$H(p \star p) - H(p) \geq g(H(p)).$$

Moreover, g is an increasing function, $\lim_{c \to \infty} g(c) = \frac{1}{8}\log_2(e)$ *and $g(c) = 0$ if and only if $c = 0$.*

Remark 1. The function g in Theorem 1 is given by

$$g(c) = \min_{x \in [0,1]} \{(cx - h_2(x)) \vee \frac{(1-x)^2((1-x) \vee (4x-2)^+)^2}{8\ln(2)}\}.$$

Remark 2. As we mentioned in the introduction, a recent result by Tao [29] implies that for a discrete $\mathbb{Z}$-valued random variable of very large entropy $H(p \star p) - H(p) \approx \frac{1}{2}$. In comparison with this result, we only get an asymptotic lower bound of $\frac{1}{8}\log_2(e) \approx 0.18$. We will see later that, the asymptotic lower bound 0.18 is also valid for independent but not necessarily identically distributed random variables provided that the entropy of both random variables approaches infinity.

The next theorem extends the i.i.d. result to the general independent case.

Theorem 2. *There is a function $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ such that for any two independent $\mathbb{Z}$-valued random variables X, X' with*

probability distributions p_1, p_2,

$$H(p_1 \star p_2) - \frac{H(p_1) + H(p_2)}{2} \geq g(H(p_1), H(p_2)).$$

Moreover, g is a positive and doubly-increasing[2] function of its arguments, $\lim_{(c,d) \to (\infty,\infty)} g(c,d) = \frac{1}{8}\log_2(e)$ *and $g(c,d) = 0$ if and only if $c = d = 0$.*

Remark 3. One might be tempted to prove the stronger bound

$$H(p_1 \star p_2) - \max\{H(p_1), H(p_2)\} \geq g(H(p_1), H(p_2)), \quad (3)$$

for some doubly-increasing function g. However, this fails because, for example, assume that p_1, p_2 are uniform distributions over $\{1, 2, \ldots, M\}$ and $\{1, 2, \ldots, NM\}$, for some number $N \geq 2$. It is not difficult to show that

$$H(p_1 \star p_2) - \max\{H(p_1), H(p_2)\} \leq \log_2(\frac{N+1}{N}),$$

which decreases to 0 with increasing N. Therefore, there is no hope to get a stronger result as in (3), which holds universally for all distributions.

The next theorem extends the results in Theorem 1 to the conditional case.

Theorem 3. *There is a function $\tilde{g} : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any two i.i.d. $\mathbb{Z}$-valued pairs of random variables (X, Y) and (X', Y'),*

$$H(X + X'|Y, Y') - H(X|Y) \geq \tilde{g}(H(X|Y)).$$

Moreover, $\tilde{g} : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function and $\tilde{g}(c) = 0$ if and only if $c = 0$.

Remark 4. The function $\tilde{g}$ is given by

$$\tilde{g}(c) = \min_{\delta \in [0, \frac{1}{2}]} \{(g(c,c) - h_2(\delta)) \vee \delta^2 g(c,c)\}, \quad (4)$$

where g is as in Theorem 2 and $h_2(\delta)$ is the binary entropy function.

III. PROOF TECHNIQUES

In this part, we will try to give an overview and also some intuition about the techniques used for proving the theorems.

A. EPI for i.i.d. random variables

We will start from the EPI for i.i.d. random variables. The main idea of the proof is to find suitable bounds for $H(p \star p) - H(p)$ in two different cases: one case in which p is a spiky distribution, namely, there is an $i \in \mathbb{Z}$ such that p_i is substantially high, and the other case where p is a quite flat and non-spiky distribution and then to combine these two bounds together.

Lemma 1. *Assume that p is a probability distribution over $\mathbb{Z}$ with $H(p) = c$ and let $x = \|p\|_\infty$. Then*

$$H(p \star p) - c \geq cx - h_2(x),$$

---

[2]A function $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ is doubly-increasing if for any value of one of the arguments, it is an increasing function of the other argument.

where h_2 is the binary entropy function.

Proof: In appendix A. ∎

Remark 5. Notice that Lemma 1, gives a very tight bound for spiky distributions for which $\|p\|_\infty$ is very close to 1, namely, for $H(p) = c$, we get $H(p \star p) - c \simeq c$, which is the best we can hope.

The next step is to give a bound for non-spiky distributions. The main idea is that in this case, it is possible to decompose the probability distribution p into two different parts p_1, p_2 with disjoint non-interlacing supports such that $p \star p_1$ and $p \star p_2$ are sufficiently far apart in ℓ_1-distance. We formalize this through the following lemmas.

Lemma 2. *Let $c > 0$, $0 < \alpha < \frac{1}{2}$ and $n \in \mathbb{Z}$. Assume that p is a probability measure over $\mathbb{Z}$ such that $\alpha \le p((-\infty, n]) \le 1 - \alpha$ and $H(p) = c$, then*

$$\|p \star p_1 - p \star p_2\|_1 \ge 2\alpha,$$

where $p_1 = \frac{1}{p((-\infty,n])} p|_{(-\infty,n]}$ and $p_2 = \frac{1}{p([n+1,\infty))} p|_{[n+1,\infty)}$ are scaled restrictions of p to $(-\infty, n]$ and $[n + 1, \infty)$ respectively.

Proof: In appendix A. ∎

Lemma 3. *Assume that p_1, p_2 and p are arbitrary probability distributions over $\mathbb{Z}$ such that p_1 and p_2 have non-overlapping supports and $\|p\|_\infty = x$. Then*

$$\|p \star p_1 - p \star p_2\|_1 \ge 2(2x - 1)^+.$$

Proof: In appendix A. ∎

Lemma 4. *Assuming the hypotheses of Lemma 2,*

$$H(p \star p) - c \ge \frac{\alpha^2}{2\ln(2)} \|p \star p_1 - p \star p_2\|_1^2.$$

Proof: In appendix A. ∎

Lemma 5. *Assume that p is a probability distribution over $\mathbb{Z}$ with $H(p) = c$ and $\|p\|_\infty = x$. Then*

$$H(p \star p) - c \ge \frac{(1 - x)^2}{8\ln(2)}((1 - x) \vee (4x - 2)^+)^2.$$

Proof: In appendix A. ∎

Now that we have the required bounds in the spiky and non-spiky cases, we can combine them to prove Theorem 1.

Proof of Theorem 1: Assume that p is a probability distribution over $\mathbb{Z}$ with $H(p) = c$ and $\|p\|_\infty = x$. It is easy to see that $x \ge 2^{-c}$. Also setting $\alpha = \frac{1-x}{2}$, there is an integer n such that $\alpha \le p((-\infty, n]) \le 1 - \alpha$. Using Lemma 1 and Lemma 5, it results that $H(p \star p) - c \ge l(c)$, where

$$l(c) = \min_{x \in [2^{-c}, 1]} \{(cx - h_2(x)) \vee \\ \frac{(1 - x)^2((1 - x) \vee (4x - 2)^+)^2}{8\ln(2)}\}.$$

We will use a simpler lower bound given by

$$g(c) = \min_{x \in [0,1]} \{(cx - h_2(x)) \vee \\ \frac{(1 - x)^2((1 - x) \vee (4x - 2)^+)^2}{8\ln(2)}\},$$

where obviously $l(c) \ge g(c)$. It is easy to check that $g(c)$ is a continuous function of c. The monotonicity of g follows from monotonicity of $cx - h_2(x)$ with respect to c, for every $x \in [0, 1]$. For strict positivity, note that $(1 - x)^2((1 - x) \vee (4x - 2)^+)^2$ is strictly positive for $x \in [0, 1)$ and it is 0 when $x = 1$, but $\lim_{x \to 1} cx - h_2(x) = c$. Hence, for $c > 0$, $g(c) > 0$. If $c = 0$ then

$$\{(cx - h_2(x)) \vee \frac{(1 - x)^2((1 - x) \vee (4x - 2)^+)^2}{8\ln(2)}\} \\ = \frac{(1 - x)^2((1 - x) \vee (4x - 2)^+)^2}{8\ln(2)},$$

and its minimum over $[0, 1]$ is 0.

For asymptotic behavior, notice that at $x = 0$, $cx - h_2(x) = 0$ and $\frac{(1-x)^2((1-x)\vee(4x-2)^+)}{8\ln(2)} = \frac{1}{8\ln(2)}$. Hence, from continuity, it results that $g(c) \le \frac{1}{8\ln(2)}$ for any $c \ge 0$. Also for any $0 < \epsilon < \frac{1}{2}$ there exists a c_0 such that for every $c > c_0$ and every x, $\epsilon < x \le 1$, $cx - h_2(x) \ge \frac{1}{8\ln(2)}$. Thus for any $\epsilon > 0$ there is a c_0 such that for $c > c_0$, the outer minimum over x in the definition of $g(c)$ is achieved on $[0, \epsilon]$, which is higher than $\frac{(1-\epsilon)^4}{8\ln(2)}$. This implies that for every $\epsilon > 0$,

$$\frac{1}{8\ln(2)} \ge \limsup_{c \to \infty} g(c) \ge \liminf_{c \to \infty} g(c) \ge \frac{(1 - \epsilon)^4}{8\ln(2)},$$

and $\lim_{c \to \infty} g(c) = \frac{1}{8\ln(2)}$. ∎

Figure 1 shows the EPI gap. As expected, the asymptotic gap is $\frac{1}{8} \log_2(e) \approx 0.18$.
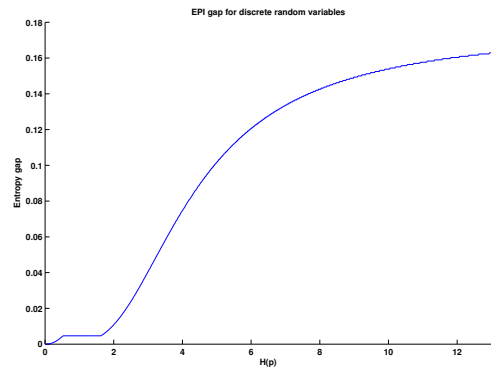

Fig. 1: The EPI gap for discrete random variables over $\mathbb{Z}$

B. EPI for non-i.i.d. random variables

Theorem 2 is an extension of Theorem 1 to independent but non identically distributed random variables. Similar to the i.i.d. case the idea is to distinguish between the spiky and non-spiky distributions.

Lemma 6. *Assume that p and q are two probability distributions over $\mathbb{Z}$ with $H(p) = c$ and $H(q) = d$. Suppose that $x = \|p\|_\infty$ and $y = \|q\|_\infty$. Then,*

$$2H(p \star q) - c - d \geq dx - h_2(x) + cy - h_2(y), \quad (5)$$

where h_2 is the binary entropy function.

    Proof: In appendix B. ∎

When at least one of the distributions is spiky, Lemma 6 gives a relatively tight bound. Hence, we should try to find a good bound for the non-spiky case.

Lemma 7. *Let p, q be two probability distributions over $\mathbb{Z}$. Assume that there are $0 < \alpha, \beta < \frac{1}{2}$ and $m, n \in \mathbb{Z}$ such that $\alpha \leq p((-\infty, m]) \leq 1 - \alpha$ and $\beta \leq q((-\infty, n]) \leq 1 - \beta$. Then*

$$\|q \star p_1 - q \star p_2\|_1 + \|p \star q_1 - p \star q_2\|_1 \geq 2(\alpha + \beta),$$

where $p_1 = \frac{1}{p((-\infty, m])} p|_{(-\infty, m]}$, $p_2 = \frac{1}{p([m+1, \infty))} p|_{[m+1, \infty)}$, $q_1 = \frac{1}{q((-\infty, n])} q|_{(-\infty, n]}$, and $q_2 = \frac{1}{q([n+1, \infty))} q|_{[n+1, \infty)}$.

    Proof: In appendix B. ∎

Lemma 8. *Assume that the hypotheses of Lemma 7 hold and let $H(p) = c$ and $H(q) = d$. Then*

$$H(p \star q) - d \geq \frac{\alpha^2}{2 \ln(2)} \|q \star p_1 - q \star p_2\|_1^2,$$

$$H(p \star q) - c \geq \frac{\beta^2}{2 \ln(2)} \|p \star q_1 - p \star q_2\|_1^2,$$

    Proof: Proof in appendix B. ∎

Lemma 9. *Let p and q be probability distributions over $\mathbb{Z}$ with $H(p) = c$, $H(q) = d$, $\|p\|_\infty = x$ and $\|q\|_\infty = y$. Then*

$$2H(p \star q) - c - d \geq l(x, y),$$

where

$$l(x, y) = \min_{(a,b) \in T(x,y)} \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8 \ln(2)},$$

and $T(x, y)$ is a subset of $(a, b) \in \mathbb{R}_+^2$ parameterized by $(x, y) \in [0, 1] \times [0, 1]$ and given by the following inequalities

$$a \geq (4y - 2)^+, b \geq (4x - 2)^+, a + b \geq 2 - x - y.$$

Moreover, $l(x, y)$ is a continuous function of (x, y), $l(x, y) \geq 0$ and $l(x, y) = 0$ if and only if $(x, y) = (1, 1)$.

    Proof: Proof in appendix B. ∎

Proof of Theorem 2: Let $x = \|p\|_\infty$ and $y = \|q\|_\infty$. It is easy to check that $x \geq 2^{-c}, y \geq 2^{-d}$. Using Lemma 6 and Lemma 9, we obtain that

$$H(p \star q) - \frac{c + d}{2} \geq s(c, d),$$

where $s(c, d)$ is given by

$$\frac{1}{2} \min_{(x,y) \in R(c,d)} \left\{ (dx - h_2(x) + cy - h_2(y)) \vee l(x, y) \right\},$$

for $R(c, d) = [2^{-c}, 1] \times [2^{-d}, 1]$. We will use a simpler lower bound given by

$$g(c, d) = \frac{1}{2} \min_{(x,y) \in R} \left\{ (dx - h_2(x) + cy - h_2(y)) \vee l(x, y) \right\},$$

where $R = [0, 1] \times [0, 1]$. It is easy to see that $g(c, d)$ is a continuous function. It is also a doubly increasing function of its arguments. To prove the last part, notice that the $l(x, y)$ in the definition of g is strictly positive except for $(x^\star, y^\star) = (1, 1)$. But $\lim_{(x,y) \to (1,1)} dx - h_2(x) + cy - h_2(y) = c + d$, which is strictly positive unless $c = d = 0$. Therefore, for $(c, d) \neq (0, 0)$, $g(c, d) > 0$.

The function $dx - h_2(x) + cy - h_2(y)$ is an increasing function of (c, d) over R, which implies that $g(c, d)$ must be an increasing function of (c, d). Also, using an argument similar to what we had in the proof of Theorem 1, it is possible to show that for high values of c and d, the outer minimum in the definition of g is achieved in a small enough neighborhood of $(0, 0)$, namely, $[0, \epsilon] \times [0, \epsilon]$ for some small enough $\epsilon > 0$. From the continuity of $l(x, y)$, it can be shown that in this range the value of $l(x, y)$ is very close to

$$\min_{(a,b): a, b \geq 0, a + b \geq 2} \frac{a^2 + b^2}{8 \ln(2)} = \frac{1}{4 \ln(2)}.$$

This implies that

$$\lim_{(c,d) \to (\infty, \infty)} g(c, d) = \frac{1}{8 \ln(2)}.$$

This completes the proof of the EPI result for the general independent case.

C. Conditional EPI

In this part, we will prove the EPI result for the conditional case, where we try to find a lower bound for the conditional entropy gap, $H(X + X'|Y, Y') - H(X|Y)$, for i.i.d. $\mathbb{Z}$-valued pairs (X, Y) and (X', Y') assuming that $H(X|Y) = c$, for some positive number c. Notice that as Y and Y' only appear in the conditioning, we do not lose generality by assuming them to be $\mathbb{Z}$-valued. Let us denote the probability distribution of Y by q then the conditional entropy gap can be written as

$$\sum_{i,j \in \mathbb{Z}} q_i q_j H(p_i \star p_j) - c,$$

where p_i is the conditional distribution of X given $Y = i$.

Notice that we are interested to the infimum of this gap over all possible q, p_i satisfying $\sum_{i \in \mathbb{Z}} q_i H(p_i) = c$. Even if the minimizing q exists, it may not be finitely supported and in general, finding the corresponding gap requires an infinite dimensional constrained optimization.

To cope with this problem, we will show that it is possible to restrict the support size of q to 2 provided that instead of the i.i.d. case we consider the general independent and non identically distributed one. Of course, at the end we get a looser bound at the price of simplifying the problem.

To be more specific, let (X, Y) and (X', Y') be independent $\mathbb{Z}$-valued pairs with $H(X|Y) = H(X'|Y') = c$ and let

$t_n(c)$ be the infimum of $H(X + X'|Y,Y') - c$ over all $(X,Y),(X',Y')$ having a conditional entropy equal to c with Y and Y' having a support size at most n. Also, assume that $t_\infty(c)$ is the corresponding infimum when there is no constraint on the support size. We first prove the following lemma.

Lemma 10. *For every $n \geq 2$, $t_\infty(c) = t_n(c)$.*

Proof: Obviously, $t_n(c) \geq t_\infty(c)$. Moreover, given any $\epsilon > 0$ there is an ϵ-optimal independent pair (X,Y) and (X',Y') such that

$$H(X + X'|Y,Y') - c \leq t_\infty(c) + \epsilon.$$

Let q, q' denote the distribution of Y, Y' and let p_i, p'_j be the conditional distribution of X, X' given $Y = i, Y' = j$. Let

$$V = \{\boldsymbol{v}_{ij} \in \mathbb{R}^3 : \boldsymbol{v}_{ij} = (H(p_i \star p'_j), H(p_i), H(p'_j)), i, j \in \mathbb{Z}\}.$$

It is easy to see that

$$\sum_{i,j \in \mathbb{Z}} q_i q'_j \boldsymbol{v}_{ij} = (H(X + X'|Y,Y'), c, c) := \boldsymbol{h},$$

which implies that the three dimensional vector $\boldsymbol{h} := (H(X + X'|Y,Y'), c, c)$ can be written as a convex combinations of the vectors $\boldsymbol{v}_{ij} \in V$ with weights $q_i q'_j$. Let $\boldsymbol{v}_i = \sum_j q'_j \boldsymbol{v}_{ij}$. Then we have $\sum_i q_i \boldsymbol{v}_i = \boldsymbol{h}$. Notice that the second component of $\boldsymbol{v}_i$ is equal to $H(p_i)$. Also, the third component is equal to c independent of i, which implies that there are only two components depending on i in $\boldsymbol{v}_i$. Therefore, by Carathèodory theorem, it is possible to write $\boldsymbol{h}$ as a convex combination of at most three $\boldsymbol{v}_i, i \in \mathbb{Z}$, which without loss of generality, we can assume to be $\{\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2\}$. In other words, there are positive $\gamma_i, i = 0, 1, 2$, $\sum_{i=0}^2 \gamma_i = 1$ and $\boldsymbol{h} = \sum_{i=0}^2 \gamma_i \boldsymbol{v}_i$. Also, note that if we change the distribution of Y from q to γ, the resulting $(X,Y), (X',Y')$ is again an ϵ-optimal solution. Now, we claim that we can simplify the problem further and find a probability triple $\psi = (\psi_0, \psi_1, \psi_2)$ with at most 2 non-zero elements such that $\sum_{i=0}^2 \psi_i H(p_i) = c$ and at the same time

$$\sum_{i=0}^2 \psi_i \boldsymbol{v}_i^{(1)} \leq \sum_{i=0}^2 \gamma_i \boldsymbol{v}_i^{(1)} = \sum_{i=0}^2 q_i \boldsymbol{v}_i^{(1)} = H(X + X'|Y,Y'),$$

where $\boldsymbol{v}_i^{(1)}$ denotes the first coordinate of the vector $\boldsymbol{v}_i$. This implies that if we replace the distribution γ for Y by ψ, which has a support of size 2, we get a lower $H(X + X'|Y,Y')$.

To prove the claim, let us consider the following optimization problem

$$\text{minimize } \sum_{i=0}^2 \psi_i \boldsymbol{v}_i^{(1)} \text{ s.t.} \begin{cases} \sum_{i=0}^2 \psi_i = 1, \\ \sum_{i=0}^2 \psi_i H(p_i) = c, \\ \psi_i \geq 0. \end{cases}$$

First of all, notice that as $\sum_{i=0}^2 \gamma_i H(p_i) = c$, γ is in the feasible set. Therefore, the feasible set is a non-empty subset of the three dimensional probability simplex. Also, as the objective function is linear in ψ, the optimal point must be

at the edge of the feasible set which implies that there is an optimal solution with at most two non-zero components and this proves the claim.

By symmetry, we can apply the same argument to the probability distribution q' of Y' to get an ϵ-optimal solution in which the support of both q and q' has at most size 2. Hence, this implies that for any $\epsilon > 0$ and any $n \geq 2$, $t_n(c) \leq t_2(c) \leq t_\infty(c) + \epsilon$. In other words, $t_n(c) = t_\infty(c)$. This completes the proof. ■

Lemma 10 allows us to simplify finding the lower bound. However, we might get a looser bound because we relaxed the condition that (X,Y) and (X',Y') be identically distributed. From now on, we will assume that Y and Y' are binary valued random variables. We will use the following two lemmas to get a lower bound for the conditional entropy gap.

Lemma 11. *Let $(X,Y), (X',Y')$ be an independent pair of random variables, where Y and Y' are binary valued with $\mathbb{P}(Y = 0) = \alpha$, $\mathbb{P}(Y' = 0) = \beta$ and $H(X|Y) = H(X' = Y') = c$. Then*

$$H(X + X'|Y,Y') - c \geq g(c,c) - \min\{h_2(\alpha), h_2(\beta)\},$$

where g is the same function as in Theorem 2.

Proof: Proof in appendix C. ■

Lemma 12. *Assume that all of the conditions of Lemma 11 hold. Suppose there is a $0 \leq \delta \leq \frac{1}{2}$ such that $\delta < \alpha, \beta < 1-\delta$. Then*

$$H(X + X'|Y,Y') - c \geq \delta^2 g(c,c).$$

Proof: Proof in appendix C. ■

Proof of Theorem 3: The proof follows by combining the results obtained in Lemma 11 and 12. Let $\delta = \min\{\alpha, 1 - \alpha, \beta, 1 - \beta\}$. Then $0 \leq \delta \leq \frac{1}{2}$ and using Lemma 12, we get the lower bound $\delta^2 g(c,c)$. Similarly, from Lemma 11 and using the fact that $\min\{h_2(\alpha), h_2(\beta)\} = h_2(\delta)$, we get the lower bound $g(c,c) - h_2(\delta)$. Combining the two, we obtain the desired lower bound

$$\tilde{g}(c) = \min_{\delta \in [0, \frac{1}{2}]} \{(g(c,c) - h_2(\delta)) \vee \delta^2 g(c,c)\}.$$

The monotonicity of $\tilde{g}$ follows from the monotonicity of $g(c,c)$. Also, notice that $\delta^2 g(c,c)$ is strictly positive unless $\delta = 0$ but $\lim_{\delta \to 0} g(c,c) - h_2(\delta) = g(c,c)$, which is strictly positive if $c > 0$. Therefore, for $c > 0$ we have $\tilde{g}(c) > 0$. This completes the proof.

IV. OPEN PROBLEMS

A. Closure convexity of the entropy set $\mathcal{H}$

As we saw in the proof of Theorem 3, the conditional EPI does not directly follow from the unconditional one. In particular, we had to relax the i.i.d. condition in order to get a relatively weak lower bound. In this part, we propose another approach to the problem which uses the closure convexity of the entropy set as we will define in a moment.

Definition 1. The entropy set $\mathcal{H}$ is defined as follows

$$\mathcal{H} := \{(H(p \star q), H(p), H(q)) \in \mathbb{R}^3_+ :$$
$$p, q \text{ are probability distributions over } \mathbb{Z}\}.$$

Remark 6. Notice that multiple (p, q) pairs may be mapped to the same point in $\mathcal{H}$ space. For example, if (p, q) is mapped to a point $\boldsymbol{v} \in \mathcal{H}$, then any distribution $(\tilde{p}, \tilde{q})$ in which $\tilde{p}$ and $\tilde{q}$ are shifted versions of p and q is also mapped to $\boldsymbol{v}$.

Remark 7. Some of the boundaries of the set $\mathcal{H}$ trivially follow from the properties of the entropy, i.e., for any $\boldsymbol{v} \in \mathcal{H}$,

$$\boldsymbol{v}^{(1)} \geq \boldsymbol{v}^{(2)}, \boldsymbol{v}^{(1)} \geq \boldsymbol{v}^{(3)},$$
$$\boldsymbol{v}^{(1)} \leq \boldsymbol{v}^{(2)} + \boldsymbol{v}^{(3)},$$

where $\boldsymbol{v}^{(i)}$ denotes the i-th coordinate of the vector $\boldsymbol{v}$. Also the boundary $\boldsymbol{v}^{(1)} = \boldsymbol{v}^{(2)} + \boldsymbol{v}^{(3)}$ is achievable. To show this, let $\boldsymbol{v}^{(2)}, \boldsymbol{v}^{(3)} \in \mathbb{R}_+$ and consider two finite support distributions p and q of support $\{0, 1, \ldots, M-1\}$ and $\{0, 1, \ldots, N-1\}$ for appropriate M and N such that $H(p) = \boldsymbol{v}^{(2)}$ and $H(q) = \boldsymbol{v}^{(3)}$. Now, fix p and define a new distribution $\tilde{q}$ as follows

$$\tilde{q}(i) = \begin{cases} 0 & \frac{i}{M} \notin \mathbb{Z}, \\ q(\frac{i}{M}) & \frac{i}{M} \in \mathbb{Z}. \end{cases}$$

It is not difficult to show that $H(\tilde{q}) = H(q) = \boldsymbol{v}^{(3)}$ and $H(p \star \tilde{q}) = H(p) + H(\tilde{q}) = \boldsymbol{v}^{(2)} + \boldsymbol{v}^{(3)}$.

We propose the following conjecture about the set $\mathcal{H}$.

Conjecture 1. *The closure of the set $\mathcal{H}$ is convex.*

Using this conjecture, we can prove the following lemma, which is a stronger version of the conditional EPI.

Theorem 4. *Assume that Conjecture 1 holds. Let (X, Y) and (X', Y') be independent pairs of $\mathbb{Z}$-valued random variables with $H(X|Y) = c, H(X'|Y') = d$. Then*

$$H(X + X'|Y, Y') - \frac{c+d}{2} \geq g(c, d),$$

where g is the same function as in Theorem 2.

Proof: Let us assume that the distribution of Y, Y' is q, q' respectively. Also assume that p_i, p'_j is the distribution of X, X' when $Y = i, Y' = j$. Let

$$\boldsymbol{v}_{ij} = (H(p_i \star p'_j), H(p_i), H(p'_j)), \; i, j \in \mathbb{Z}.$$

Notice that $\boldsymbol{v}_{ij} \in \mathcal{H}$. We also have

$$(H(X + X'|Y, Y'), c, d) = \sum_{i,j \in \mathbb{Z}} q_i q'_j \boldsymbol{v}_{ij},$$

which is a convex combination of the vectors $\boldsymbol{v}_{ij}$. By the closure convexity of $\mathcal{H}$, for any $\epsilon > 0$ it is possible to find an $\boldsymbol{h} \in \mathcal{H}$ in ϵ-neighborhood of $(H(X + X'|Y, Y'), c, d)$. In other words, for the given $\epsilon > 0$, there are two distributions μ_1, μ_2 over $\mathbb{Z}$ such that

$$H(\mu_1 \star \mu_2) - \epsilon \leq H(X + X'|Y, Y') \leq H(\mu_1 \star \mu_2) + \epsilon,$$
$$H(\mu_1) - \epsilon \leq c \leq H(\mu_1) + \epsilon,$$
$$H(\mu_2) - \epsilon \leq d \leq H(\mu_2) + \epsilon.$$

In particular, this implies that

$$H(X + X'|Y, Y') - \frac{c+d}{2}$$
$$\geq H(\mu_1 \star \mu_2) - \frac{c+d}{2} - \epsilon$$
$$\geq H(\mu_1 \star \mu_2) - \frac{H(\mu_1) + H(\mu_2)}{2} - 2\epsilon$$
$$\geq g(H(\mu_1), H(\mu_2)) - 2\epsilon$$
$$\geq g(c - \epsilon, d - \epsilon) - 2\epsilon,$$

where we used the monotonicity of g with respect to both arguments. As $\epsilon > 0$ is arbitrary and g is a continuous function, it results that $H(X + X'|Y, Y') - \frac{c+d}{2} \geq g(c, d)$. ∎

Remark 8. In the case that (X, Y) and (X', Y') are i.i.d. pairs with $H(X|Y) = H(X'|Y') = c$, this result reduces to

$$H(X + X'|Y, Y') - c \geq g(c, c),$$

which is tighter than the bound (4) obtained in Theorem 3.

REFERENCES

[1] C. Shannon,"A mathematical theory of communications, I and II," Bell Systems Technical Journal, vol. 27, pp. 379423, 1948.

[2] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," Information and Control, vol. 2, no. 2, pp. 101112, 1959.

[3] N. Blachman, "The convolution inequality for entropy powers," IEEE Transactions on Information Theory, vol. 11, no. 2, pp. 267–271, 1965.

[4] E. Lieb, "Proof of an entropy conjecture of Wehrl," Communications in Mathematical Physics, vol. 62, no. 1, pp. 3541, 1978.

[5] M. Costa, "A new entropy power inequality," IEEE Transactions on Information Theory, vol. 31, no. 6, pp. 751–760, 1985.

[6] S. Verdù and D. Guo, "A simple proof of the entropy power inequality," IEEE Transactions on Information Theory, vol. 52, no. 5, pp. 2165–2166, 2006.

[7] O. Rioul, "Information theoretic proofs of entropy power inequalities," IEEE Transactions on Information Theory, vol. 57, no. 1, pp. 33–55, 2011.

[8] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," IEEE Transactions on Information Theory, vol. 19, no. 2, pp. 197–207, 1973.

[9] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," IEEE Transactions on Information Theory, vol. 24, no. 4, pp. 451–456, 1978.

[10] L. Ozarow, "On a source-coding problem with two channels and three receivers," Bell Syst. Tech. J, vol. 59, no. 10, pp. 1909–1921, 1980.

[11] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 1057–1070, 1998.

[12] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 3936–3964, 2006.

[13] M. Costa, "A new entropy power inequality," IEEE Transactions on Information Theory, vol. 31, no. 6, pp. 751760, 1985.

[14] A. Dembo, Simple proof of the concavity of the entropy power with respect to added Gaussian noise, IEEE Transactions on Information Theory, vol. 35, no. 4, pp. 887–888, 1989.

[15] C. Villani, A short proof of the concavity of entropy power," IEEE Transactions on Information Theory, vol. 46, no. 4, pp. 1695–1696, 2000.

[16] R. Zamir and M. Feder, "A generalization of the entropy power inequality with applications," IEEE Transactions on Information Theory, vol. 39, no. 5, pp. 1723–1728, 1993.

[17] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," IEEE Transactions on Information Theory, vol. 53, no. 5, pp. 1839–1851, 2007.

[18] R. Liu, T. Liu, H. Poor, and S. Shamai, "A vector generalization of Costas entropy-power inequality with applications," IEEE Transactions on Information Theory, vol. 56, no. 4, pp. 1865–1879, 2010.

[19] S. Artstein, K. Ball, F. Barthe, and A. Naor, "Solution of Shannons problem on the monotonicity of entropy," Journal of the American Mathematical Society, vol. 17, no. 4, pp. 975–982, 2004.

[20] A. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof," IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 4295–4297, 2006.

[21] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," IEEE Transactions on Information Theory, vol. 53, no. 7, pp. 2317–2329, 2007.

[22] S. Haghighatshoar, E. Abbe, E. Telatar, "Adaptive sensing using deterministic partial Hadamard matrices," In Proc. International Symposium on Information Theory, pp. 1842–1846, 2012.

[23] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications I," IEEE Transactions on Information Theory, vol. 19, no. 6, pp. 769772, 1973.

[24] H. Witsenhausen, "Entropy inequalities for discrete channels," IEEE Transactions on Information Theory, vol. 20, no. 5, pp. 610616, 1974.

[25] P. Harremoes, C. Vignat, "An entropy power inequality for the binomial family," Journal of Inequalities in Pure and Applied Mathematics, vol. 4, no. 5, 2003.

[26] N. Sharma, S. Das, and S. Muthukrishnan, "Entropy power inequality for a family of dis- crete random variables," In Proc. of International Symposium on Information Theory, pp. 1945–1949, 2011.

[27] O. Johnson, Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," IEEE Transaction on Information Theory, vol. 56, pp. 5387– 5395, 2010.

[28] A. Jog, V. Anantharam, "The Entropy Power Inequality and Mrs. Gerber's Lemma for Abelian Groups of Order 2^n," arXiv:1207.6355, 2012.

[29] T. Tao,"Sumset and inverse sumset theory for Shannon entropy," Combinatorics, Probability & Computing, vol. 19, no. 4, pp. 603639, 2010.

[30] I. Kontoyiannis and M. Madiman, "Sumset and inverse sumset inequalities for differential entropy and mutual information," arXiv:1206.0489, 2012.

APPENDIX A
EPI FOR I.I.D. RANDOM VARIABLES

Proof of Lemma 1: Assume that X is a $\mathbb{Z}$-valued random variable with probability distribution p. Let $i \in \mathbb{Z}$ be such that $p(i) = \|p\|_\infty = x$. Let p_i be the probability distribution p shifted by i, i.e., $p_i(k) = p(k - i)$ for every $k \in \mathbb{Z}$. Assume that $P := p_i$. Note that $H(p \star p) = H(P \star P)$ and $H(P) = H(p) = c$. Let B be a binary random variable with $\mathbb{P}\{B = 0\} = x = 1 - \mathbb{P}\{B = 1\}$, and let R be a random variable defined by $\mathbb{P}\{R = k\} = p_i(k)/(1 - x)$ for every $k \in \mathbb{Z} \setminus \{0\}$ and $\mathbb{P}\{R = 0\} = 0$. Note that $X = BR$ for independent B and R. We also have $H(X) = h_2(x) + (1 - x)H(R)$. Let X' be an independent copy of X. Then, we have

$$
\begin{aligned}
H(P \star P) &= H(BR + X') \\
&\geq H(BR + X'|B) \\
&= xc + (1 - x)H(X' + R) \\
&\geq xc + (1 - x)H(R) \\
&= xc + c - h_2(x).
\end{aligned}
$$

This yields $H(p \star p) - c \geq xc - h_2(x)$. ∎

Proof of Lemma 2: Let $\alpha_1 = p((-\infty, n])$ and $\alpha_2 = p([n + 1, \infty)) = 1 - \alpha_1$. Note that $p = \alpha_1 p_1 + \alpha_2 p_2$. We distinguish two cases $\alpha_1 \leq \frac{1}{2}$ and $\alpha_1 > \frac{1}{2}$. If $\alpha_1 \leq \frac{1}{2}$ then we have

$$
\begin{aligned}
&\|p \star p_1 - p \star p_2\| \\
&= \|\alpha_1 p_1 \star p_1 - (1 - \alpha_1)p_2 \star p_2 + (1 - 2\alpha_1)p_1 \star p_2\|_1 \\
&\geq \|\alpha_1 p_1 \star p_1 - (1 - \alpha_1)p_2 \star p_2\|_1 - (1 - 2\alpha_1)\|p_1 \star p_2\|_1 \\
&= \alpha_1 + (1 - \alpha_1) - (1 - 2\alpha_1) = 2\alpha_1 \geq 2\alpha,
\end{aligned}
$$

whereas if $\alpha_1 > \frac{1}{2}$ we have

$$
\begin{aligned}
&\|p \star p_1 - p \star p_2\| \\
&= \|\alpha_1 p_1 \star p_1 - (1 - \alpha_1)p_2 \star p_2 + (1 - 2\alpha_1)p_1 \star p_2\|_1 \\
&\geq \|\alpha_1 p_1 \star p_1 - (1 - \alpha_1)p_2 \star p_2\|_1 - (2\alpha_1 - 1)\|p_1 \star p_2\|_1 \\
&= \alpha_1 + (1 - \alpha_1) - (2\alpha_1 - 1) = 2(1 - \alpha_1) \geq 2\alpha,
\end{aligned}
$$

where we used the triangle inequality, $1 - \alpha_1 \geq \alpha$ and the fact that $p_1 \star p_1$ and $p_2 \star p_2$ have non-overlapping supports, so the ℓ_1-norm of the sum is equal to sum of the corresponding ℓ_1-norms. ∎

Proof of Lemma 3: Let $n_0 \in \mathbb{Z}$ be such that $p(n_0) = \|p\|_\infty = x$. We have

$$
\begin{aligned}
\|p \star p_1 - p \star p_2\|_1 &= \sum_{i \in \mathbb{Z}} |p \star p_1(i) - p \star p_2(i)| \\
&= \sum_{i \in \mathbb{Z}} |\sum_{j \in \mathbb{Z}} p(j)(p_1(i - j) - p_2(i - j))| \\
&\geq \sum_{i \in \mathbb{Z}} p(n_0)|p_1(i - n_0) - p_2(i - n_0)| \\
&\quad - \sum_{i \in \mathbb{Z}} \sum_{j \neq n_0} p(j)|p_1(i - j) - p_2(i - j)| \\
&= x\|p_1 - p_2\|_1 - (1 - x)\|p_1 - p_2\|_1 \\
&= 2(2x - 1),
\end{aligned}
$$

where we used the fact that p_1 and p_2 have non-overlapping supports thus $\|p_1 - p_2\|_1 = \|p_1\|_1 + \|p_2\|_1 = 2$. As $\|p \star p_1 - p \star p_2\|_1 \geq 0$, we have $\|p \star p_1 - p \star p_2\|_1 \geq 2(2x - 1)^+$. ∎

Proof of Lemma 4: Let α_1 and α_2 be the same as in the proof of Lemma 2. Let $\nu_1 = p_1 \star p$, $\nu_2 = p_2 \star p$, and for $x \in [0, 1]$, define $\mu_x = x\nu_1 + (1 - x)\nu_2$ and $f(x) = H(\mu_x)$. We have

$$
\begin{aligned}
f'(x) &= -\sum (\nu_{1i} - \nu_{2i}) \log_2(\mu_{xi}), \\
f''(x) &= -\frac{1}{\ln(2)} \sum \frac{(\nu_{1i} - \nu_{2i})^2}{\mu_{xi}} \leq 0.
\end{aligned}
$$

Therefore, $f(x)$ is a concave function of x. Moreover,

$$
\begin{aligned}
f'(0) &= D(\nu_1\|\nu_2) + H(\nu_1) - H(\nu_2), \\
f'(1) &= -D(\nu_2\|\nu_1) + H(\nu_1) - H(\nu_2).
\end{aligned}
$$

Since p_1 and p_2 have different supports, there are i, j such that $\nu_{1i} = 0, \nu_{2i} > 0$ and $\nu_{1j} > 0, \nu_{2j} = 0$. Hence $D(\nu_1\|\nu_2)$ and $D(\nu_2\|\nu_1)$ are both equal to infinity. In other words,

$$
f'(0) = +\infty, f'(1) = -\infty.
$$

Hence, the unique maximum of the function f must happen between 0 and 1. Assume that for fixed ν_1 and ν_2, $x^\star$ is the maximizer. If $0 < \alpha_1 \le x^\star$ then

$$\alpha_1 f^{'}(\alpha_1) = \sum \alpha_1 (\nu_{2i} - \nu_{1i}) \log_2(\mu_{\alpha_1 i}) \ge 0,$$

which implies that

$$\begin{aligned}
f(\alpha_1) &= -\sum \mu_{\alpha_1 i} \log_2(\mu_{\alpha_1 i}) \\
&= -\sum (\nu_{2i} + \alpha_1(\nu_{1i} - \nu_{2i})) \log_2(\mu_{\alpha_1 i}) \\
&\ge -\sum \nu_{2i} \log_2(\mu_{\alpha_1 i}) \\
&= H(\nu_2) + D(\nu_2 \| \mu_{\alpha_1}) \\
&\ge H(p) + \frac{1}{2\ln(2)} \| \nu_2 - \mu_{\alpha_1} \|_1^2 \\
&= H(p) + \frac{\alpha_1^2}{2\ln(2)} \| \nu_1 - \nu_2 \|_1^2,
\end{aligned}$$

where we used Pinsker's inequality for distributions r and s,

$$D(r \| s) \ge \frac{1}{2\ln(2)} \| r - s \|_1^2.$$

Similarly, we can show that if $x^\star \le \alpha_1 \le 1$ then

$$f(\alpha_1) \ge H(p) + \frac{(1-\alpha_1)^2}{2\ln(2)} \| \nu_1 - \nu_2 \|_1^2.$$

As $\alpha \le \alpha_1 \le 1 - \alpha$ and $\alpha \le \frac{1}{2}$ it results that

$$\begin{aligned}
H(p \star p) &= H(\alpha_1 p \star p_1 + (1 - \alpha_1) p \star p_2) \\
&= f(\alpha_1) \\
&\ge H(p) + \frac{\alpha^2}{2\ln(2)} \| \nu_1 - \nu_2 \|_1^2 \\
&\ge c + \frac{\alpha^2}{2\ln(2)} \| \nu_1 - \nu_2 \|_1^2. \qquad \blacksquare
\end{aligned}$$

Proof of Lemma 5: Let $x = \|p\|_\infty$ and $\alpha = \frac{1-x}{2}$. It is easy to show that there is an $n \in \mathbb{Z}$ such that $\alpha \le p((-\infty, n]) \le 1 - \alpha$. Also let p_1 and p_2, as in Lemma 2, be the restriction of p to $(-\infty, n]$ and $[n+1, \infty)$. As p_1 and p_2 have disjoint supports, using Lemma 2 and 3, it results that

$$\| p \star p_1 - p \star p_2 \|_1 \ge (1 - x) \vee (4x - 2)^+,$$

Therefore, using Lemma 4, we get

$$H(p \star p) - c \ge \frac{(1-x)^2}{8\ln(2)}((1-x) \vee (4x-2)^+)^2. \qquad \blacksquare$$

APPENDIX B
EPI FOR NON-I.I.D. RANDOM VARIABLES

Proof of Lemma 6: Let X and Y be two independent random variables with probability distribution p and q. Similar to the proof of Lemma 1, there is a binary random variable B, $\mathbb{P}(B = 0) = x$ and a random variable R independent of B such that $\tilde{X} = BR$, where $\tilde{X}$ is a suitably shifted version of X

such that $\mathbb{P}(\tilde{X} = 0) = x$. Also, $H(X) = h_2(x) + (1-x)H(R)$. Then, we get

$$\begin{aligned}
H(p \star q) &= H(X + Y) \\
&= H(\tilde{X} + Y) = H(BR + Y) \\
&\ge H(BR + Y | B) \\
&\ge \mathbb{P}(B = 0)H(Y) + \mathbb{P}(B = 1)H(R + Y) \\
&\ge xd + (1-x)H(R) \\
&= xd + c - h_2(x),
\end{aligned}$$

which implies that $H(p \star q) - c \ge xd - h_2(x)$. By symmetry, we also obtain that $H(p \star q) - d \ge yc - h_2(y)$. Combining these two results we get

$$2H(p \star q) - c - d \ge dx - h_2(x) + cy - h_2(y). \qquad \blacksquare$$

Proof of Lemma 7: Let $\alpha_1 = p((-\infty, m])$, $\alpha_2 = 1 - \alpha_1$, $\beta_1 = q((-\infty, n])$ and $\beta_2 = 1 - \beta_1$. Note that $p = \alpha_1 p_1 + \alpha_2 p_2$ and $q = \beta_1 q_1 + \beta_2 q_2$. Thus we obtain

$$\begin{aligned}
&\| q \star p_1 - q \star p_2 \|_1 + \| p \star q_1 - p \star q_2 \|_1 \\
&\ge \| q \star p_1 - q \star p_2 + p \star q_1 - p \star q_2 \|_1 \\
&= \| (\alpha_1 + \beta_1) p_1 \star q_1 + (\beta_2 - \alpha_1) p_1 \star q_2 \\
&\quad + (\alpha_2 - \beta_1) p_2 \star q_1 - (\alpha_2 + \beta_2) p_2 \star q_2 \|_1 \\
&\ge \| (\alpha_1 + \beta_1) p_1 \star q_1 - (\alpha_2 + \beta_2) p_2 \star q_2 \|_1 \\
&\quad - \| (\beta_2 - \alpha_1) p_1 \star q_2 + (\alpha_2 - \beta_1) p_2 \star q_1 \|_1 \\
&\ge \alpha_1 + \beta_1 + \alpha_2 + \beta_2 - |\beta_2 - \alpha_1| - |\alpha_2 - \beta_1| \\
&= 2(1 - |1 - (\alpha_1 + \beta_1)|),
\end{aligned}$$

where we used the triangle inequality and the fact that $p_1 \star q_1$ and $p_2 \star q_2$ have non-overlapping supports. Now, two cases can happen: if $\alpha_1 + \beta_1 \le 1$ then $(1 - |1 - (\alpha_1 + \beta_1)|) = (\alpha_1 + \beta_1) \ge (\alpha + \beta)$. Otherwise, $\alpha_1 + \beta_1 > 1$ and we obtain

$$\begin{aligned}
(1 - |1 - (\alpha_1 + \beta_1)|) &= 2 - (\alpha_1 + \beta_1) \\
&= \alpha_2 + \beta_2 \ge \alpha + \beta.
\end{aligned}$$

Therefore, in both cases we get

$$\| q \star p_1 - q \star p_2 \|_1 + \| p \star q_1 - p \star q_2 \|_1 \ge 2(\alpha + \beta),$$

which is the desired result. $\qquad \blacksquare$

Proof of Lemma 8: Let $\alpha_1 := p((-\infty, m])$, $\alpha_2 := 1 - \alpha_1$, $\nu_1 := p_1 \star q$, $\nu_2 := p_2 \star q$, and for $x \in [0, 1]$, let $\mu_x := x\nu_1 + (1-x)\nu_2$ and $f(x) := H(\mu_x)$. By an argument similar to what we had in the proof of Lemma 4, we can show that

$$H(p \star q) = f(\alpha_1) \ge d + \frac{\alpha^2}{2\ln(2)} \| \nu_1 - \nu_2 \|_1^2,$$

which implies that

$$H(p \star q) - d \ge \frac{\alpha^2}{2\ln(2)} \| q \star p_1 - q \star p_2 \|_1^2.$$

The other inequality in the lemma follows by symmetry. $\qquad \blacksquare$

Proof of Lemma 9: As $\|p\|_\infty = x, \|q\|_\infty = y$, setting $\alpha = \frac{1-x}{2}$ and $\beta = \frac{1-y}{2}$ and using Lemma 8, we obtain

$$2H(p \star q) - c - d \geq \frac{\alpha^2 a^2 + \beta^2 b^2}{2\ln(2)}$$
$$= \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8\ln(2)}$$

where $a = \|q \star p_1 - q \star p_2\|_1$ and $b = \|p \star q_1 - p \star q_2\|_1$. Also, from Lemma 7, we have

$$a + b \geq 2(\alpha + \beta) = 2 - x - y. \tag{6}$$

Furthermore, applying Lemma 3 to the distribution p with $\|p\|_\infty = x$ and q_1, q_2 with disjoint supports, and similarly to q with $\|q\|_\infty = y$ and p_1, p_2 with disjoint supports, we get

$$b \geq (4x - 2)^+, a \geq (4y - 2)^+. \tag{7}$$

Therefore,

$$2H(p \star q) - c - d \geq l(x, y),$$

where

$$l(x, y) = \min_{(a,b) \in T(x,y)} \frac{(1-x)^2 a^2 + (1-y)^2 b^2}{8\ln(2)},$$

and $T(x, y)$ is defined by the three inequalities derived in (6) and (7).

The continuity of $l(x, y)$ can be easily checked. For the last part of the lemma, notice that if $M := x \vee y < 1$ then it is not difficult to show that

$$l(x, y) \geq \min_{a+b \geq 2-2M} \frac{(1-M)^2}{8\ln(2)} (a^2 + b^2) \geq \frac{(1-M)^4}{4\ln(2)} > 0,$$

which is strictly positive. Moreover, if $x \vee y = 1$ but $(x, y) \neq (1, 1)$ then, for example, $y \in [0, 1), x = 1$, which implies that $b \geq 2$. Therefore, we get $l(x, y) \geq \frac{(1-y)^2}{2\ln(2)}$, which is strictly positive unless $y = 1$. A similar argument applies to $x \in [0, 1), y = 1$. Therefore, over $(x, y) \in [0, 1] \times [0, 1]$, $l(x, y) \geq 0$ and $l(x, y) = 0$ if and only if $(x, y) = (1, 1)$. ∎

APPENDIX C
CONDITIONAL EPI

Proof of Lemma 11: To prove the lemma, notice that we have the constraint $H(X|Y) = H(X'|Y') = c$ and the probability distribution of Y, Y' has a support of size 2. We first prove that it is possible to modify the conditional distribution of the random variables X and X' given Y and Y' in a way that none of the constraints are violated, $H(X + X'|Y, Y')$ remains fixed and simultaneously, $H(Y|X)$ and $H(Y'|X')$ become as small as we want. To show this , let p_i, p'_j, $i, j \in \{0, 1\}$ be the distribution of X, X' conditioned on $Y = i, Y' = j$. Notice that if we shift any p_i, p'_j to the right or to the left by as many steps as we want, the conditional entropies remain unchanged so does $H(X + X'|Y, Y')$. We claim that by suitable shift of distributions, it is possible to make $H(Y|X)$ as small as we want. The same is true for $H(Y'|X')$.

To prove the claim, let $\epsilon > 0$ and assume that A_ϵ and B_ϵ are subsets of $\mathbb{Z}$ of minimal size such that $p_0(A_\epsilon) \geq 1 - \epsilon/2$ and $p_1(B_\epsilon) \geq 1 - \epsilon/2$. In particular, for any $i \in A_\epsilon, j \in B_\epsilon$, $p_0(i) > 0, p_1(j) > 0$. Moreover,

$$\mathbb{P}(X \in A_\epsilon \cup B_\epsilon) \geq \alpha p_0(A_\epsilon) + (1 - \alpha) p_1(B_\epsilon)$$
$$\geq 1 - \frac{\epsilon}{2}.$$

For $n \in \mathbb{Z}_+$, let us define $B_\epsilon^{(n)} = \{i + n : i \in B_\epsilon\}$, to be the right shift of B_ϵ by n. Also assume that $p_1^{(n)}$ is the probability distribution shifted to the right by n, namely, for $k \in \mathbb{Z}$, $p_1^{(n)}(k) = p_1(k - n)$. Specially, this implies that

$$p_1^{(n)}(B_\epsilon^{(n)}) = p_1(B_\epsilon).$$

Now let us replace p_1, by $p_1^{(n)}$ and let us the denote the resulting random variable by $\tilde{X}$. This assumption does not change $H(X|Y)$ and $H(X + X'|Y, Y')$. As A_ϵ and B_ϵ are finite sets, there is N_1 such that for all $n > N_1$, the two sets A_ϵ and $B_\epsilon^{(n)}$ are disjoint. For $a \in A_\epsilon$ and $b \in B_\epsilon$, let us compute the conditional distribution of Y given $\tilde{X} = a$ and $\tilde{X} = b + n \in B_\epsilon^{(n)}$. We have

$$\mathbb{P}(Y = 0|\tilde{X} = a) = \frac{\alpha p_0(a)}{\alpha p_0(a) + (1 - \alpha) p_1(a - n)},$$
$$\mathbb{P}(Y = 1|\tilde{X} = b + n) = \frac{(1 - \alpha) p_1(b)}{(1 - \alpha) p_1(b) + \alpha p_0(b + n)}.$$

It is not difficult to see that for all $a \in A_\epsilon$ and all $b \in B_\epsilon$, both of these numbers converge to 1 as n goes to infinity which implies that both $H(Y|\tilde{X} = a)$ and $H(Y|\tilde{X} = b)$ converge to 0. In particular, there is an N_2 such that for $n > N_2$ these two numbers are less than $\frac{\epsilon}{2}$. Therefore, for $n > \max\{N_1, N_2\}$ we have

$$H_n(Y|\tilde{X}) = \sum_{k \in \mathbb{Z}} p_{\tilde{X}}(k) H(Y|\tilde{X} = k)$$
$$\leq \sum_{k \in A_\epsilon \cup B_\epsilon^{(n)}} p_{\tilde{X}}(k) \times \frac{\epsilon}{2} + \sum_{k \notin A_\epsilon \cup B_\epsilon^{(n)}} p_{\tilde{X}}(k) \times 1$$
$$= \sum_{k \in A_\epsilon \cup B_\epsilon} p_X(k) \times \frac{\epsilon}{2} + \sum_{k \notin A_\epsilon \cup B_\epsilon} p_X(k) \leq \epsilon,$$

which proves the claim. Now assume that we have selected $(X, Y), (X', Y')$ such that $H(Y|X), H(Y'|X') < \epsilon$ for some positive small number ϵ. Then we have

$$H(X + X'|Y, Y') - c$$
$$= H(X + X') - H(X) - I(X + X'|Y, Y') + I(X; Y)$$
$$\geq H(X + X') - H(X) - H(Y, Y') + H(Y) - H(Y|X)$$
$$\geq H(X + X') - H(X) - H(Y, Y') + H(Y) - \epsilon$$
$$\geq H(X + X') - H(X) - H(Y') - \epsilon$$
$$\geq g(H(X), H(X')) - h_2(\beta) - \epsilon$$
$$\geq g(c, c) - h_2(\beta) - \epsilon,$$

where we used the independence of Y, Y', increasing property of g and the fact that $H(X) \geq H(X|Y) = c$ and similarly

$H(X') \geq c$. As this is true for any $\epsilon > 0$, we obtain

$$H(X + X'|Y, Y') - c \geq g(c, c) - h_2(\beta).$$

By symmetry, we also have

$$H(X + X'|Y, Y') - c \geq g(c, c) - h_2(\alpha).$$

Therefore, we get the desired result

$$H(X + X'|Y, Y') - c \geq g(c, c) - \min\{h_2(\alpha), h_2(\beta)\}.$$

Proof of Lemma 12: Assuming the hypotheses of Lemma 11, there must be $i, j \in \{0, 1\}$ such that $H(p_i), H(p'_j) \geq c$. Therefore, we have

$$
\begin{aligned}
H(X + &X'|Y, Y') - c \\
&= \sum_{k,l=0}^{1} q_k q'_l \left(H(p_k \star p'_l) - \frac{H(p_k) + H(p'_l)}{2} \right) \\
&\geq q_i q'_j \left(H(p_i \star p'_j) - \frac{H(p_i) + H(p'_j)}{2} \right) \\
&\geq \delta^2 g(c, c).
\end{aligned}
$$